**Answers:**

**1. Where did you get the dataset, and how did you ensure its reliability?**

* We obtained the dataset from **Figshare**, a publicly available data repository.
* To ensure reliability:
  + We checked for **data completeness** and consistency.
  + We reviewed **existing literature** to confirm the relevance of the dataset.
  + We validated the data using **statistical analysis** to check for anomalies.

**2. What preprocessing steps did you perform, and why?**

* **Handled missing values**:
  + **Categorical columns** (e.g., location, state) → Filled with the **mode** (most frequent value).
  + **Numerical columns** (e.g., precipitation, projected yield) → Filled with the **median**, as the data was skewed.
* **Outlier treatment**:
  + Used the **Interquartile Range (IQR) method** to remove extreme outliers.
* **Standardization & Normalization**:
  + Standardized numerical features to improve model performance.
* **Feature Encoding**:
  + Categorical variables were converted into **numerical values** for ML models.

**3. Did you remove any features? If so, why?**

* Yes, we removed some **irrelevant or redundant features**:
  + **Highly correlated features** → Removed to prevent multicollinearity.
  + **Features with too many missing values** → If a column had more than **40% missing data**, it was dropped.
  + **ID-like columns** (e.g., unique identifiers) → These do not contribute to predictions.

**4. What were the key insights from your EDA?**

* **Temperature and precipitation trends** significantly impact crop yields.
* **Maize was the most vulnerable crop** to climate change, showing the highest number of negative outliers.
* **Rice had the most stable yields**, indicating resilience to climate fluctuations.
* **Geospatial analysis** showed that regions near the **equator** experienced stronger negative climate impacts.

**5. Why did you choose specific visualizations?**

* **Pairplots**: To identify correlations between **climate factors and yield**.
* **Boxplots**: To detect and analyze **outliers** in the dataset.
* **Geospatial maps**: To show the **regional distribution** of climate change impacts.
* **Time series plots**: To analyze **historical trends** in crop yields.

**6. Did you notice any unexpected trends or anomalies in the data?**

* Yes, some anomalies were observed:
  + **Extreme outliers in precipitation data**, likely due to measurement errors or rare climate events.
  + Some crops showed **unexpected resilience** to climate change in certain regions.
  + A few **negative yield values**, which were corrected during preprocessing.

**7. Why did you choose the specific machine learning models (e.g., Random Forest, XGBoost)?**

* **Random Forest Regression**: Handles **non-linear relationships**, is robust to **outliers**, and prevents **overfitting**.
* **XGBoost**: More accurate than traditional models due to **gradient boosting**, **feature importance ranking**, and **handling missing data efficiently**.
* **Logistic Regression (for classification tasks)**: Acts as a **baseline model** for identifying high-risk areas.
* **Neural Networks**: Considered for capturing **complex interactions** but required more computational resources.

**8. How did you handle categorical variables in your dataset?**

* **One-Hot Encoding** for nominal categories (e.g., crop type, region).
* **Label Encoding** for ordinal categories if applicable.

**9. Did you consider any other algorithms or approaches? If so, why did you reject them?**

* **Linear Regression**: Too simplistic; failed to capture **non-linear effects**.
* **Decision Trees**: Overfitted to the training data.
* **Support Vector Machines (SVMs)**: Computationally expensive for large datasets.

**10. What were the most important features in predicting crop yields?**

* **Temperature and precipitation levels** had the highest influence.
* **CO₂ concentration** had a secondary but notable impact.
* **Soil quality and regional location** played a significant role in yield variation.

**11. What do your results imply for future crop yields under different climate scenarios?**

* **Crops will become more vulnerable** in certain regions if temperature and precipitation trends continue.
* **Adapting agricultural practices** (e.g., better irrigation, climate-resistant crops) is necessary.
* Some **regions will require major policy interventions** to prevent food shortages.

**12. What were the biggest challenges you faced during the project?**

* **Data Cleaning**: Handling missing values and outliers required careful imputation.
* **Computational Complexity**: Advanced models like **XGBoost and Neural Networks** required high processing power.
* **Data Imbalance**: Some crops had **much more data** than others, requiring balancing techniques.

**13. Are there any limitations in your analysis or dataset?**

* **Limited geographic scope**: The dataset may not cover **all climate zones globally**.
* **Data sparsity**: Some crop types had fewer records, leading to potential bias.
* **Assumption of linear impact**: The models assume that climate factors **impact yields in a predictable way**, which may not always hold true.

**14. How can your findings be applied in real-world agriculture or policy-making?**

* **Farmers** can use the predictions to **adjust planting times and irrigation strategies**.
* **Policymakers** can design **climate adaptation programs** for vulnerable regions.
* **Agricultural companies** can plan **investment in climate-resilient crops**.

**15. How does your work contribute to the broader field of climate change research?**

* Demonstrates **how machine learning can predict climate impacts** on agriculture.
* Helps **governments and policymakers prepare for food security challenges**.
* Encourages **further research** in AI-driven climate impact modeling.

**16. How did you handle missing data in your dataset?**

* Used **mode imputation** for categorical features.
* Used **median imputation** for numerical features with skewed distributions.
* Verified **no major loss of data quality** after imputations.

**17. Why did you choose this project?**

* Understanding **climate change’s impact on global food supply** is crucial for:
  + **Food security**.
  + **Sustainable agriculture**.
  + **Developing AI-driven solutions** for climate adaptation.